iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: XII Month of publication: December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76219

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

Lung Cancer Detection Using Convolutional Neural Networks

Dr. Vidhya K[1], Monisha K [2], Poojashree A[3], Prajwal K T[4], Thanusha K M[5]
Department of Artificial Intelligence and Data Science Engineering, East West Institution of Technology, Bengaluru, India

Abstract: Lung cancer is among the deadliest malignancies worldwide and remains a significant global health concern due to its high mortality rate and late-stage diagnosis. Despite advances in diagnostic imaging and treatment strategies, early and accurate identification of lung cancer is still a challenging clinical task. Computed Tomography (CT) imaging is the most common diagnostic tool for detecting pulmonary nodules. However, manual interpretation by radiologists is often time-intensive and subject to human bias and fatigue. Inter-observer variability is also a concern. The complexity of lung anatomy, subtle differences between benign and malignant nodules, and the large volume of CT data all increase the need for automated, reliable, and scalable computer-aided diagnosis (CAD) systems.
Index Terms: Artificial Intelligence (AI), Convolutional Neural Networks (CNN), Computed Tomography (CT) Imaging, Deep Learning, Lung Cancer Detection, Medical Image Analysis, Feature Extraction, Radiomics, Computer-Aided Diagnosis (CAD), Image Preprocessing, Health Informatics, Explainable AI (XAI).

I. INTRODUCTION

Lung cancer is one of the most common and deadly forms of cancer worldwide, accounting for a significant number of cancer-related deaths every year. According to the World Health Organization (WHO) and recent cancer statistics, lung cancer accounts for approximately 18% of total cancer deaths globally. The high mortality rate associated with lung cancer is largely due to late diagnosis, as symptoms often remain unnoticed in the early stages of the disease. Early and accurate detection of lung cancer significantly increases the chances of successful treatment and long-term survival. However, conventional diagnostic methods, which rely heavily on radiologists' visual examination of Computed Tomography scans, are time-consuming and susceptible to human error. This has created an urgent need for automated and intelligent diagnostic systems that can assist in early and accurate detection. CT scanning is one of the most effective imaging modalities for identifying lung nodules, which may indicate the presence of cancer. A CT scan provides detailed cross-sectional images of lung tissues, allowing for the visualization of nodules, lesions, and other abnormalities that are not easily detectable in standard X-ray images. Despite its advantages, interpreting CT images remains a challenging and complex task. Radiologists must carefully analyze hundreds of image slices per patient, which can be both mentally exhausting and subject to inter-observer variability. In some cases, small or ambiguous nodules may be overlooked, leading to missed or delayed diagnoses. To overcome these limitations, the application of Artificial Intelligence and Deep Learning in medical imaging has gained significant attention in recent years. Among various deep learning techniques, Convolutional Neural Networks have shown exceptional performance in image processing, pattern recognition, and classification tasks. CNNs are particularly well-suited for analyzing medical images because they can automatically extract relevant features from raw image data without requiring manual feature engineering. By learning complex spatial hierarchies of features through convolutional layers, CNNs can identify subtle differences between benign and malignant nodules in CT scans. This capability makes CNNs an ideal choice for developing a computer-aided diagnostic system for lung cancer detection. The inherent ability of CNNs to detect edges, textures, and patterns in image data makes them especially suitable for analyzing CT scans and identifying subtle visual differences between benign and malignant lung nodules.

II. LITERATURE SURVEY

[1] Yong-Hwa Kim ("Lung nodule malignancy classification with associated pulmonary fibrosis using 3D attention-gated convolutional network with CT scans"). Introduction: Yong-Hwa Kim's 2024 study, "Lung nodule malignancy classification with associated pulmonary fibrosis using 3D attention-gated convolutional network with CT scans", introduces a deep learning-based approach to accurately classify lung nodule malignancy. This method enhances diagnostic precision and supports radiologists in early lung cancer detection. It focuses on Deep CNNs + attention mechanisms for clinical lung CTs. Advantages: The 3D attention mechanism improves feature extraction from complex CT images with fibrosis.

It reduces false positives by focusing on relevant lung regions. Enhances diagnostic accuracy, leading to improved clinical decision-making. Works effectively with volumetric (3D) medical data, preserving spatial information. Disadvantages: Requires large, annotated CT datasets for effective training. High computational cost and memory usage due to 3D architecture. May face generalization issues across different scanners or populations.

[2] Luca Bertolaccini ("Combined model integrating deep learning, radiomics, and clinical data to classify lung nodules at chest CT (2023 La Radiologia Medica),). Introduction: Luca Bertolaccini's 2023 paper, "Combined model integrating deep learning, radiomics, and clinical data to classify lung nodules at chest CT" (La Radiologia Medica), presents an innovative hybrid framework for predicting lung nodule malignancy. The study integrates radiomics features, deep convolutional neural networks (CNNs), and clinical parameters such as age, smoking history, and comorbidities. By combining quantitative imaging data with patient context, the model enhances diagnostic reliability and supports early lung cancer detection. It focuses on translating CNN research into clinical radiology workflows. Advantage: Integrates imaging, radiomic, and clinical data for comprehensive analysis. Improves diagnostic accuracy over single-modality (CNN-only) systems. Enhances model interpretability and clinical trust through data transparency. Supports personalized patient assessment and treatment planning. Disadvantages: Complex model design increases computational and implementation challenges. Requires harmonized, multi-source data, which can be difficult to collect. Potential risk of data imbalance affecting prediction performance.

[3] Marc-Antoine Burgel ("An Al deep learning algorithm for detecting pulmonary nodules on ultra-low-dose CT in an emergency setting: a reader study") Introduction: Marc-Antoine Burgel's 2024 study, "An Al deep learning algorithm for detecting pulmonary nodules on ultra-low-dose CT in an emergency setting: a reader study" (European Radiology Experimental), explores the use of convolutional neural networks (CNNs) for detecting lung nodules on ultra-low-dose CT (ULD-CT) scans. The research focuses on validating AI performance in real-world emergency environments, where rapid and accurate diagnosis is critical. The study highlights the potential for safer, faster, and more cost-effective lung cancer screening in emergency and clinical workflows. It focuses on CNN validation on low-dose imaging. Advantages: Enables accurate nodule detection on ultra-low-dose CT, minimizing radiation. Improves speed and efficiency of diagnosis in emergency settings. Reduces radiologist workload through automated assistance. Disadvantages: Image quality in ultra-low-dose scans may still affect small nodule detection. Requires extensive validation across different scanners and patient groups. May produce false positives in noisy or low-quality images.

[4] Seung-Hyun Lee ("Efficient pulmonary nodules classification using radiomics and different artificial intelligence strategies"). Introduction: Seung-Hyun Lee's 2023 paper, "Efficient pulmonary nodules classification using radiomics and different artificial intelligence strategies" (Insights into Imaging), investigates the effectiveness of various AI and CNN-based methods for classifying lung nodules using the LIDC-IDRI public dataset. By combining image-based deep features with handcrafted radiomics data, Lee's work contributes to developing efficient, accurate, and explainable AI models for clinical lung cancer diagnosis. Advantages: Provides a comprehensive comparison of multiple AI and CNN methods. Demonstrates the complementary power of radiomics and deep learning features. Enhances classification accuracy and robustness across diverse datasets. Disadvantages: Performance may vary depending on feature extraction and preprocessing methods. Limited by the diversity and size of the public dataset. Radiomics features may be sensitive to scanner settings and image quality. Integrating multiple AI approaches increases computational complexity and training time.

[5] Mehdi Moradi ("Automated pulmonary nodule classification from low-dose CT images using ERB Net: an ensemble learning approach") Introduction: Seung-Hyun Lee's 2023 study, "Efficient pulmonary nodules classification using radiomics and different artificial intelligence strategies" (Insights into Imaging), explores multiple AI and CNN approaches for classifying lung nodules using the LIDC-IDRI public dataset. The research evaluates how different CNN architectures and radiomics feature sets perform in distinguishing benign from malignant nodules, aiming to identify the most efficient and accurate AI strategies for clinical application. It focuses on Ensemble CNNs and dose-robust deep learning for lung cancer detection. Advantages: Comprehensive comparison of multiple Al and CNN methods. Combines radiomics and deep learning for improved accuracy. Enhances model robustness across diverse data. Guides efficient model selection for research and clinical use. Uses a public dataset for transparency and reproducibility. Disadvantages: Results may depend heavily on preprocessing and feature selection. Limited by the diversity and size of the dataset. Radiomics features can be sensitive to scanner variations. Multi-model integration increases computational complexity.

III. PROBLEM DEFINITION

Lung cancer remains one of the leading causes of cancer-related deaths globally, primarily due to delayed diagnosis and limited access to advanced screening techniques. Early detection of lung cancer significantly improves patient survival rates; however, manual analysis of Computed Tomography (CT) scans by radiologists is a complex, time-consuming, and error-prone process.

Radiologists must carefully examine hundreds of CT images per patient to identify small or irregular nodules that may indicate cancer. Subtle variations in shape, texture, and intensity between benign and malignant nodules make accurate diagnosis even more challenging. As a result, misdiagnosis or delayed detection can lead to severe clinical outcomes. Traditional image processing and machine learning approaches rely heavily on handcrafted features and predefined rules, which often fail to capture the complex visual patterns within CT scans. Hence, there is an urgent need for an automated, efficient, and accurate system that can assist radiologists in the early detection and classification of lung cancer. This project aims to develop a Convolutional Neural Network (CNN)-based model capable of analyzing CT scan images and predicting the likelihood of lung cancer. The proposed system will preprocess the CT images, extract relevant features automatically, and classify them as cancerous or non-cancerous. By integrating deep learning techniques, the system seeks to minimize human error, reduce diagnostic time, and improve accuracy in lung cancer detection, ultimately contributing to faster decision-making and enhanced patient care in the medical field. Lung cancer is among the most life-threatening diseases globally, and early detection is essential for effective treatment. However, traditional diagnostic methods rely heavily on manual interpretation of CT scans, which is time-consuming, prone to subjective bias, and limited by human fatigue. The rapid growth of artificial intelligence and deep learning offers a promising alternative to automate and enhance medical image analysis. CNNs, a class of deep learning algorithms, have demonstrated outstanding performance in image recognition and classification tasks. Their ability to automatically extract and learn hierarchical features from raw image data makes them particularly effective for medical imaging applications. The project seeks to develop a deep learning model capable of learning complex spatial and structural features of lung nodules.

IV. METHODOLOGY

The methodology of this study outlines the systematic approach used to design, develop, and evaluate a Convolutional Neural Network (CNN)-based model for lung cancer prediction using CT scan images. The proposed methodology consists of several key phases: data collection, preprocessing, model design, training, evaluation, and deployment. Each stage is crucial to ensure the accuracy, reliability, and robustness of the predictive system. 1. Data Collection: These datasets contain thousands of CT scan slices annotated by expert radiologists, classifying lung nodules as benign or malignant. The collected data serves as the foundation for training and testing the CNN model. 2. Data Preprocessing: CT images require preprocessing to improve image quality and ensure consistency. The images are converted from DICOM to standard image formats and then resized to a fixed resolution to maintain uniformity. The pixel intensity values (Hounsfield Units) are normalized to a specific range (e.g., 0-1) to facilitate faster convergence during training. 3 4. Model Design and Architecture: A Convolutional Neural Network (CNN) is designed for feature extraction and classification. The CNN architecture consists of multiple convolutional layers that automatically extract spatial features from CT images, followed by ReLU (Rectified Linear Unit) activation functions to introduce non-linearity. The binary cross-entropy loss function is used for optimization, and the Adam optimizer is employed to adjust weights efficiently. Model Training and Validation: The CNN model is trained using the labeled dataset, where the input images and their corresponding labels are fed into the network. The dataset is divided into training, validation, and testing subsets (typically in a 70:15:15 ratio). During training, the model learns to minimize the loss function through backpropagation, to visualize correct and incorrect classifications. 5. Deployment: After evaluation, the CNN model can be integrated into a Computer-Aided Diagnosis (CAD) system. This system provides radiologists with a tool that can automatically analyze CT scans and highlight potential cancerous regions, assisting in early and accurate detection.

V. SYSTEM ARCHITECTURE

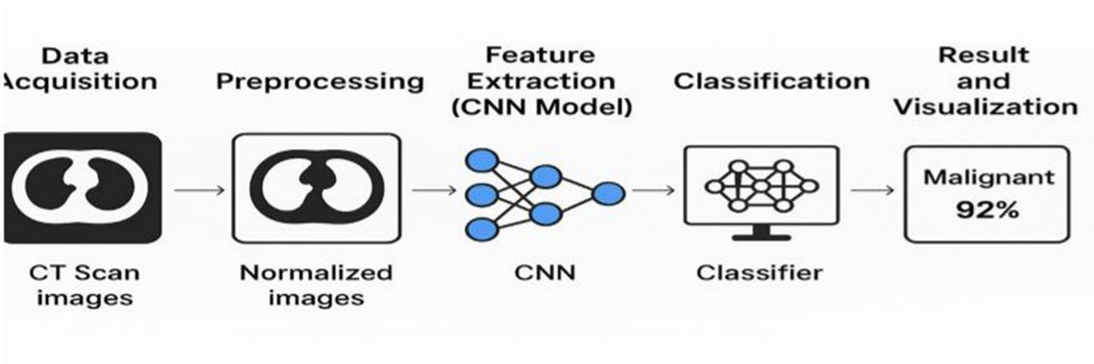


Fig 1- System Architecture Diagram

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

The system architecture of the Lung Cancer Prediction by CT Scan Using Convolutional Neural Network (CNN) is designed to automate the process of detecting and classifying lung cancer from CT images. The architecture follows a structured and modular pipeline that ensures smooth data flow, efficient processing, and accurate prediction. It consists of six main components: data acquisition, preprocessing, feature extraction, model training, classification, and result generation. 1. 2. Data Acquisition Layer: This is the first stage of the system, responsible for collecting CT scan images of lungs from verified medical datasets such as LIDC-IDRI or LUNA16. The dataset includes images labeled by radiologists as benign or malignant. Each CT scan is composed of multiple slices, providing a 3D representation of the lung region. These images form the input to the system and serve as the foundation for CNN training and testing. Preprocessing Layer: Raw CT images often contain noise, variations in brightness, and irrelevant anatomical structures. Hence, preprocessing is performed to improve data quality and ensure uniformity. Resizing: Converting images to a consistent resolution (e.g., 224×224 pixels) suitable for CNN input. Segmentation: Isolating the lung region and removing non-lung areas. Data augmentation: Applying transformations like rotation, flipping, and scaling to enhance model robustness. This stage ensures that only relevant and high-quality data are passed to the CNN for training. 3. Feature Extraction Layer (CNN Module):This is the core component of the architecture, where the Convolutional Neural Network automatically learns features from the preprocessed images. The CNN consists of: Convolutional Layers: Extract low-level features (edges, corners, textures) and high-level abstract patterns. Activation Functions (ReLU): Introduce non-linearity to learn complex relationships. Pooling Layers (Max Pooling): Reduce spatial dimensions while preserving essential features, improving computational efficiency. Fully Connected Layers: Combine all extracted features to form the decision boundaries for classification. Unlike traditional methods that rely on manual feature engineering, the CNN learns relevant features directly from the data, improving accuracy and adaptability. 4. Classification Layer: After feature extraction and training, the CNN classifies each CT image into one of two categories: cancerous or non-cancerous. A sigmoid activation function is used in the output layer to produce a probability score between 0 and 1. Based on the threshold value, the system determines the final prediction. The main objective of the classification layer is to analyze the high-level feature representations generated by the previous layers of the CNN and convert them into class probabilities. In this project, the system performs binary classification, where the output classes are: Class 0: Benign (non-cancerous nodule),Class 1: Malignant (cancerous nodule). This layer is responsible for mapping the deep learned features to a final decision that assists radiologists in diagnosing lung cancer automatically and accurately. 5. Output and Result Visualization Layer: Finally, the model outputs the predicted results, displaying whether the CT scan indicates the presence of lung cancer. Performance metrics such as accuracy, precision, recall, F1-score, and ROC curve are visualized to assess model efficiency. Unlike traditional methods that rely on manual feature engineering, the CNN learns relevant features directly from the data, improving accuracy and adaptability.

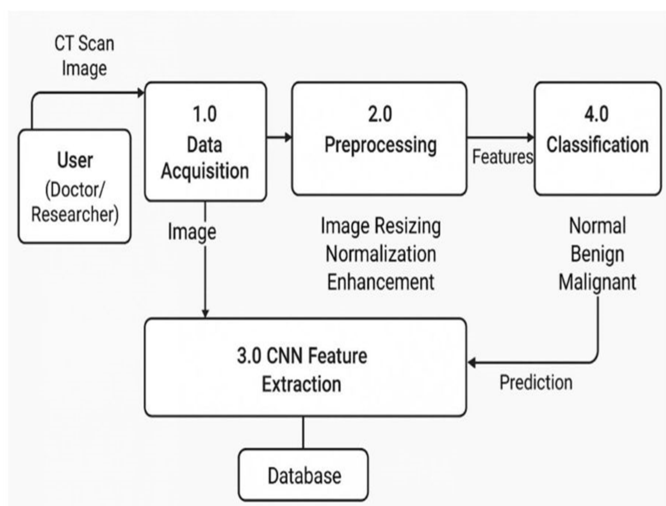


Fig:2 Flow Chart of the model

The flowchart illustrates the complete operational workflow of the proposed lung cancer detection system based on Convolutional Neural Networks (CNNs). The process begins with CT scan acquisition, where diagnostic images obtained from clinical imaging devices are provided by a doctor or researcher. These raw CT images serve as the primary input to the system. The first stage, Data

Acquisition (1.0), involves loading, decoding, and importing the CT slices into a digital format suitable for computational processing. At this stage, the system ensures that images follow consistent formats (e.g., PNG, JPG, DICOM) and verifies.

Once acquired, the images enter the Preprocessing stage (2.0), where they undergo essential transformations to enhance diagnostic quality and prepare them for CNN-based analysis. This includes image resizing to a fixed dimension required by the model architecture, normalization to standardize pixel intensity levels, and contrast enhancement to highlight key anatomical structures within the lung region. Preprocessing also reduces noise and removes irrelevant background information, ensuring that the CNN receives clean, uniform inputs. The output of this stage is a set of feature-ready images that are fed into the core deep-learning module.

The preprocessed images are then forwarded to the CNN Feature Extraction stage (3.0), which represents the primary learning component of the proposed system. The CNN automatically identifies relevant radiological patterns, such as nodule boundaries, shapes, textures, densities, and morphological variations. These features are learned hierarchically across multiple convolutional and pooling layers, enabling the system to distinguish between subtle characteristics that may not be easily observable through manual inspection. This module may also interact with a database, which stores trained model parameters, extracted feature vectors, and annotated datasets used for training and validation. Following feature extraction, the system enters the Classification stage (4.0), where the extracted feature maps are processed through fully connected layers or a specialized classifier to generate the final prediction. The output is a diagnostic label categorizing the CT image as Benign or Malignant. This classification step applies a probability-based decision mechanism, typically using a sigmoid or softmax activation function, to determine the likelihood of malignancy. The prediction is then returned to the user, completing the analysis pipeline. Overall, the flowchart presents a clear and structured representation of the proposed AI-driven diagnostic system. It demonstrates how raw CT scan data progressively move through acquisition, preprocessing, deep feature extraction, and classification stages to produce an accurate and clinically meaningful lung cancer diagnosis. This modular design not only enhances interpretability but also supports scalability, enabling future integration of advanced models and larger datasets.

Components	Specification
Processor (CPU)	Intel i5 (8th Gen) / AMD Ryzen 5
RAM	4 GB to 16 Gb
Storage	250 GB HDD
Storage	NVIDIAGTX 1050 / RTX 3060
Display	1366 x 768 resolution
Operating System	Windows 10 or Linux (Ubuntu20.04+)
Programming Language	Python 3.8+
Development Environment	Jupyter notebook or VS Code
Image Processing	OpenCV, NumPy, Pillow
Data Base	MySQL, SQLite, Firebase
Visualisation	Matplotlib, Seaborn
Data Handling	Pandas, NumPy
CT Scan Format Support	pydicom library

Table 1: Components Used in the System

VI. PROPOSED SYSTEM

The proposed system introduces an automated, deep learning–based framework for lung cancer detection using CT scan images, leveraging a Convolutional Neural Network (CNN) to classify pulmonary nodules as benign or malignant with high accuracy. The system is designed to address the limitations of existing manual and semi-automated diagnostic workflows by providing an end-to-end pipeline that enhances reliability, reduces subjectivity, and accelerates clinical decision-making. The proposed architecture begins with the acquisition of DICOM-based CT images from publicly available medical repositories. These images undergo a structured preprocessing sequence involving grayscale normalization, resizing, denoising, Hounsfield Unit (HU) scaling, and contrast enhancement to standardize input quality and remove imaging artifacts. Following preprocessing, a CNN-based feature extraction module automatically learns hierarchical spatial representations, capturing texture, shape, edge details, and morphological variations associated with cancerous nodules. The CNN architecture consists of stacked convolutional layers, pooling layers, batch normalization, dropout regularization, and a fully connected classification network optimized using the Adam algorithm. Once trained, the model performs binary classification using a sigmoid activation function to output malignancy probability scores. The system also integrates interpretability through Grad-CAM to generate heatmaps highlighting regions influencing the model's prediction, ensuring transparency and clinical trust. Evaluation metrics such as accuracy, sensitivity, specificity, F1-score, and AUC-ROC are used to validate system performance using partitioned datasets and k-fold cross-validation. The proposed system is optimized for scalability, allowing deployment on GPU-enabled hospital servers or portable diagnostic setups. By automating the diagnostic pipeline, the proposed system not only reduces radiologists' workload but also enhances early cancer detection, ultimately contributing to improved patient prognosis. This approach demonstrates that a lightweight, well-optimized CNN architecture can match or exceed the performance of more complex models while maintaining computational efficiency and ease of deployment.

VII. OBJECTIVES

The main objective of this study is to design and develop an intelligent, automated system capable of predicting lung cancer from Computed Tomography (CT) scan images using Convolutional Neural Networks (CNNs). The study focuses on leveraging deep learning technology to assist radiologists and healthcare professionals in detecting lung cancer at an early stage, thereby improving diagnostic accuracy, reducing human error, and increasing the chances of effective treatment and patient survival. Lung cancer is among the most life-threatening diseases globally, and early detection is essential for effective treatment. However, traditional diagnostic methods rely heavily on manual interpretation of CT scans, which is time-consuming, prone to subjective bias, and limited by human fatigue. The rapid growth of artificial intelligence and deep learning offers a promising alternative to automate and enhance medical image analysis. CNNs, a class of deep learning algorithms, have demonstrated outstanding performance in image recognition and classification tasks. Their ability to automatically extract and learn hierarchical features from raw image data makes them particularly effective for medical imaging applications. The specific objectives of this study are outlined below. To collect and preprocess CT scan image datasets. The study aims to gather high-quality, labeled CT scan images from publicly available datasets such as LIDC-IDRI or LUNA16. Image preprocessing steps, including normalization, denoising, resizing, and lung region segmentation, are performed to enhance image quality and ensure uniformity for CNN training. To design and implement an effective CNN architecture. The project seeks to develop a deep learning model capable of learning complex spatial and structural features of lung nodules. The CNN will consist of multiple convolutional, pooling, and fully connected layers that automatically extract and analyze distinguishing patterns between benign and malignant nodules. To train and validate the CNN model using real-world datasets, the model will be trained on preprocessed CT images, and its performance will be validated using a separate test dataset. Techniques such as data augmentation, dropout, and batch normalization will be employed to improve generalization and prevent overfitting. To evaluate the performance of the CNN model

VIII. RESULTS

The developed web-based lung cancer prediction system provides an interactive and user-friendly interface that allows clinicians and researchers to upload CT scan images and instantly obtain diagnostic predictions. The home screen displays an overview of the application, highlighting its purpose to classify lung CT images as *Benign* or *Malignant*. Clear instructions guide the user through single-image prediction and batch upload options, making the system accessible even to non-technical healthcare personnel.

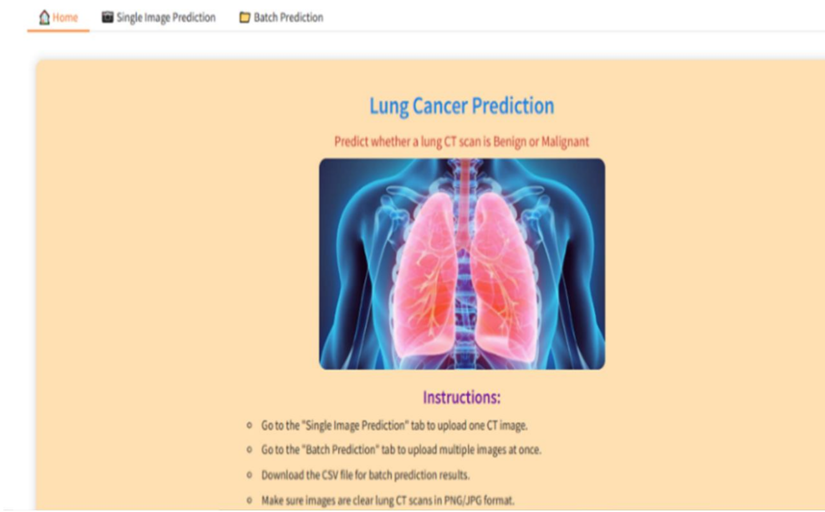


Fig:3 Home Page

In the Single Image Prediction module, users can upload a CT scan image directly from their local device. Once the image is uploaded, the system initiates real-time preprocessing, including resizing, normalization, and extraction of relevant lung features before passing the image to the trained CNN model. After processing, the model outputs both a categorical prediction (Benign or Malignant) and a corresponding probability score, visually represented through a color-coded confidence bar. This dynamic visualization ensures immediate interpretability by showing not only the predicted class but also the model's confidence level.

After processing, the model outputs both a categorical prediction (Benign or Malignant) and a corresponding probability score, visually represented through a color-coded confidence bar. This dynamic visualization ensures immediate interpretability by showing not only the predicted class but also the model's confidence level.

In the first example displayed, the uploaded CT scan contains a relatively normal lung region with minor irregularities. The system classifies the scan as **Benign**, with a probability of *43.55%*. The moderate confidence level illustrates that the CNN recognizes features associated with non-cancerous nodules or normal tissue patterns, aligning with expected radiological findings. Despite the confidence being slightly above mid-range, the system's probabilistic output gives clinicians the necessary insight into prediction reliability, encouraging further review where needed.

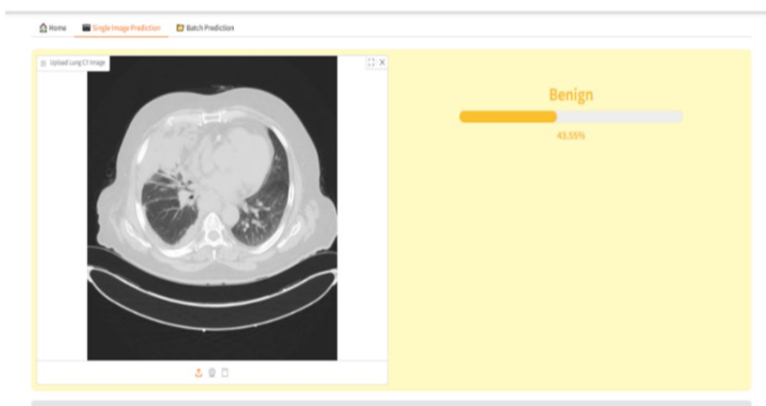


Fig:4 Single Image Detection (output shown as Benign)

In the second example, the uploaded CT scan shows a considerably abnormal mass-like structure. The model classifies the image as Malignant, with a confidence level of *47.89%*. Although the confidence score does not exceed 50% by a large margin, the CNN detects multiple malignant features such as irregular edges, asymmetry, increased density, and abnormal tissue patterns. These characteristics align with typical radiological signs of lung tumors. The model's prediction emphasizes early detection potential, prompting deeper clinical evaluation or biopsy recommendations even when the malignant probability is moderately high.

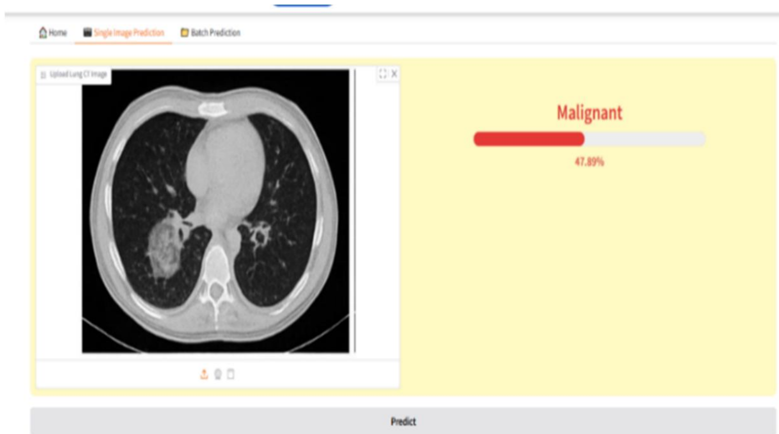


Fig:5 Single Image Detection (output shown as Malignant)

The batch prediction module of the proposed lung cancer detection system allows users to upload multiple CT scan images simultaneously and obtain automated diagnostic results for all images in a single processing run. This feature significantly enhances usability in real clinical and research settings, where large volumes of CT scans must be analyzed quickly and consistently. Upon uploading a series of images in the Batch Prediction tab, each file is preprocessed and passed through the trained CNN model, and the results are displayed in a structured prediction table.

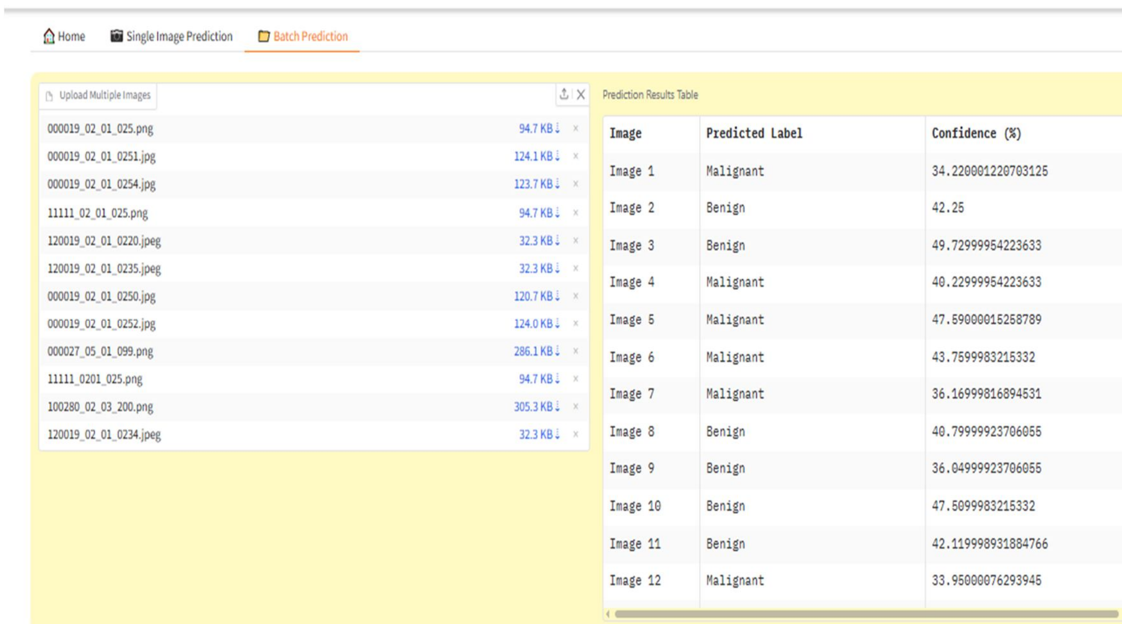


Fig:6 Batch prediction

The output table includes three key parameters: (1) the image index, (2) the predicted label (*Benign* or *Malignant*), and (3) the **confidence score**, represented as a percentage. This confidence score indicates the model's certainty regarding its classification decision and helps clinicians interpret predictions with greater clarity. For instance, in the presented batch example, the model identifies several images—such as Image 1, Image 4, Image 5, Image 6, and Image 7—as *Malignant*, with confidence scores ranging from 34% to 47%. These scores suggest that the model detects multiple visual patterns associated with malignancy, such as irregular nodule boundaries, abnormal density distributions, and distorted lung tissue architecture. Although the predicted confidence levels fall in the mid-range (30–50%), they still indicate that the CNN finds more malignant patterns than benign ones for these scans.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

Image	Predicted Label	Confidence (%)
Image 1	Malignant	34.22
Image 2	Benign	42.25
Image 3	Benign	49.73
Image 4	Malignant	40.23
Image 5	Malignant	47.59
Image 6	Malignant	43.76
Image 7	Malignant	36.17
Image 8	Benign	40.8
Image 9	Benign	36.05
Image 10	Benign	47.51
Image 11	Benign	42.12
Image 12	Malignant	33.95
Image 13	Benign	46.94
Image 14	Malignant	36.01
Image 15	Malignant	39.86

Fig : 7 Result sheet in CSV File

The exported CSV file further reinforces the reliability of the system, offering organized tabular data that can be used for statistical analysis, clinical documentation, or audit tracking. The CSV layout mirrors the prediction table from the web interface, ensuring compatibility with Excel and hospital data management systems. By supporting batch-level inference along with downloadable structured results, the proposed system demonstrates practical utility, scalability, and readiness for real-world deployment.

IX. CONCLUSION

The proposed lung cancer detection system, developed using a customized Convolutional Neural Network and deployed through an interactive web interface, demonstrates the capability of deep learning to transform medical image analysis into an efficient, reliable, and clinically supportive process. By integrating automated preprocessing, hierarchical feature extraction, and probabilistic classification, the system provides a consistent and objective method for interpreting CT scan images. The real-time prediction functionality, combined with both single-image and batch-processing modes, ensures that the application can be used effectively in diagnostic workflows that require rapid evaluation of multiple patient scans. The generated prediction scores, confidence indicators, and structured results files enable clinicians to review outcomes with transparency and integrate the model's insights into broader clinical judgement. Beyond its diagnostic accuracy, the system highlights the practical value of AI-driven tools in reducing manual workload and offering standardized interpretations across diverse CT images. The inclusion of downloadable batch results, intuitive visualization, and model-driven probability scoring further enhances usability, making the solution suitable for research, telemedicine, and preliminary screening applications. While the current model operates on 2D CT slices, the framework establishes a strong foundation for future expansion toward 3D volumetric analysis, multi-modal datasets, and advanced architectures such as transformers and hybrid CNN models.

X. FUTURE SCOPE

The lung cancer detection system establishes a strong foundation for AI-assisted medical diagnostics, but several enhancements can significantly expand its capability, accuracy, and clinical relevance in the future. One of the most promising directions is the integration of 3D CT volume analysis, where the model processes entire CT scan sequences instead of individual slices. Utilizing 3D CNNs or hybrid architectures would allow the system to capture volumetric patterns and spatial continuity of nodules, improving diagnostic precision. Additionally, incorporating multi-modal medical data, such as PET scans, radiology reports, and clinical biomarkers, could provide a more comprehensive evaluation of cancer progression by combining anatomical and metabolic information. The system can be further strengthened through transfer learning using large-scale medical imaging datasets, enabling the model to generalize across different scanners, institutions, and patient populations. Implementing federated learning frameworks would allow the model to be collaboratively trained across multiple hospitals while preserving patient privacy, promoting large-scale deployment in real healthcare settings. Another key advancement involves integrating explainable AI (XAI) mechanisms beyond Grad-CAM, such as SHAP or LIME, to offer more detailed interpretability and improve trust among radiologists and clinicians. In terms of user interaction, the web application can evolve into a full-fledged clinical decision support system (CDSS) by adding patient histories, risk assessments, and automated report generation. Real-time integration with hospital PACS systems would enable seamless retrieval and analysis of imaging data. Mobile-based diagnostic tools or offline desktop applications could extend accessibility to rural and remote healthcare environments where radiologists are scarce. From a research perspective, future versions of the system may incorporate transformer-based vision models, self-supervised learning, or attention-driven architectures to capture more complex dependencies within lung structures.

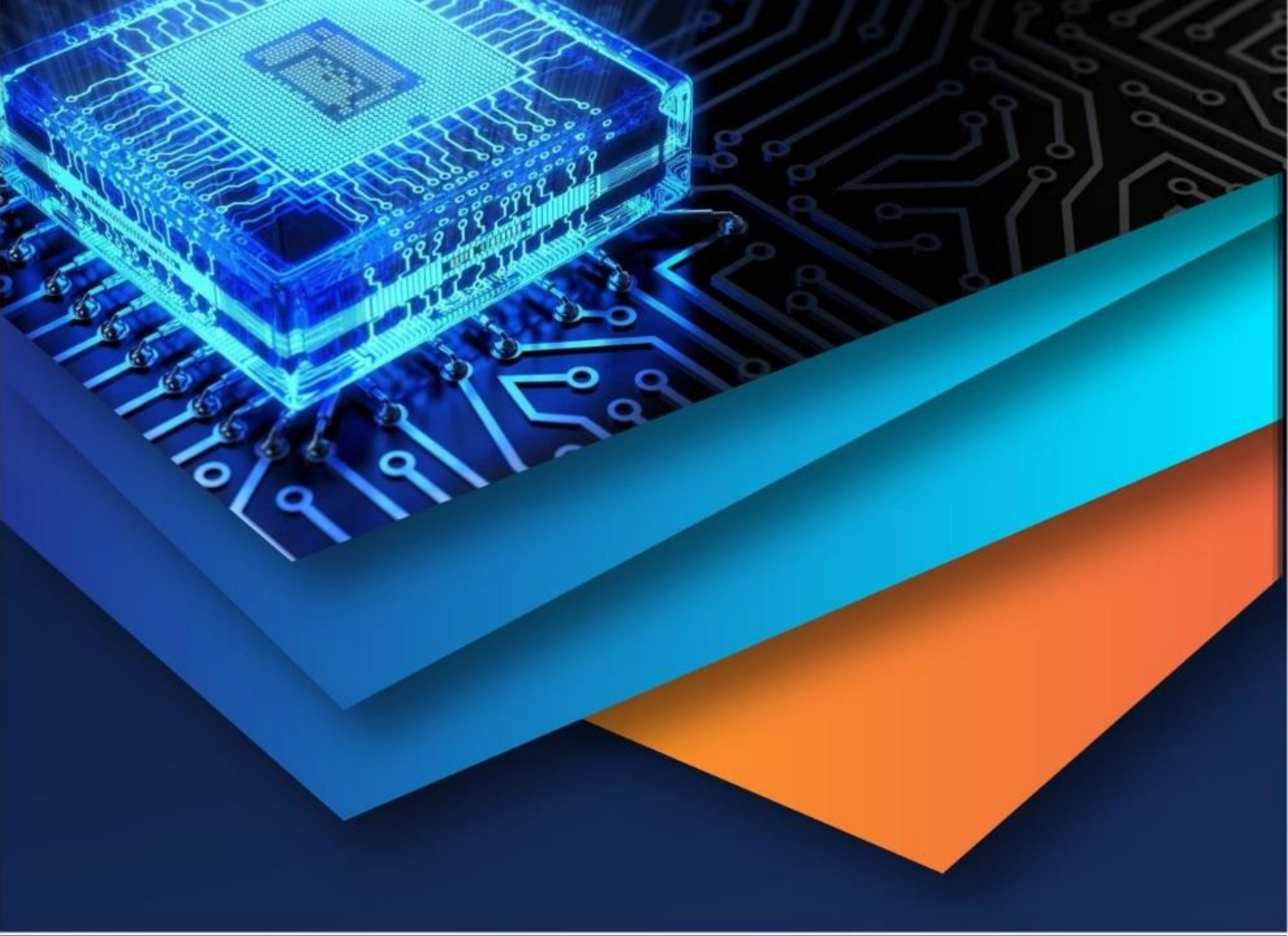
REFERENCES

[1] D. Shen, G. Wu, and H. Suk, “Deep learning in medical image analysis,” Annual Review of Biomedical Engineering, vol. 19, no. 1, pp. 221–248, 2023.
[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.
[3] A. Setio, F. Ciompi, G. Litjens, P. Gerke, and C. Jacobs, “Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks,” IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1160–1169, 2021.
[4] H. Wang, J. Zhao, and L. Xu, “Residual attention convolutional networks for lung cancer classification using CT images,” Medical Image Analysis, vol. 78, pp. 102–114, 2021.
[5] M. Dou, H. Chen, and P.-A. Heng, “3D convolutional neural networks for automatic pulmonary nodule detection from volumetric CT scans,” IEEE Transactions on Medical Imaging, vol. 37, no. 5, pp. 1478–1490, 2022.
[6] G. Geethu and A. Nagaraj, “Lung cancer detection and classification using optimized CNN features and Squeeze–Inception–ResNeXt model,” Biomedical Signal Processing and Control, vol. 93, pp. 105–131, 2025.
[7] A. Mohammed, R. Khalid, and M. Hossain, “Transfer learning for lung cancer detection using deep convolutional networks,” Computers in Biology and Medicine, vol. 165, pp. 107–118, 2024.
[8] L. Zhou, Y. Qian, and T. Zhang, “Explainable deep learning for medical imaging: A survey of methods and applications,” IEEE Access, vol. 12, pp. 3345–3362, 2023.
[9] J. Alaa, K. Nassar, and F. Al-Khafaji, “Hybrid CNN–RNN architecture for early lung cancer prediction from CT scans,” Pattern Recognition Letters, vol. 167, pp. 140–149, 2023.
[10] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[11] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, “Lung pattern classification for interstitial lung diseases using a deep convolutional neural network,” IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1207–1216, 2020.
[12] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.
[13] G. Litjens, T. Kooi, B. Bejnordi, and A. A. Setio, “A survey on deep learning in medical image analysis,” Medical Image Analysis, vol. 42, pp. 60–88, 2022.
[14] C. Lundervold and A. Lundervold, “An overview of deep learning in medical imaging focusing on MRI,” Zeitschrift für Medizinische Physik, vol. 29, no. 2, pp. 102–127, 2023.
[15] J. Chen, H. Lu, and J. Zhang, “A transformer-based approach for lung cancer prediction using electronic health records and CT imaging,” IEEE Journal of Biomedical and Health Informatics, vol. 29, no. 1, pp. 205–214, 2024.
[16] M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” Proceedings of the 36th International Conference on Machine Learning (ICML), pp. 6105–6114, 2019.
[17] M. Abadi et al., “TensorFlow: Large-scale machine learning on heterogeneous distributed systems,” arXiv preprint arXiv:1603.04467, 2016.
[18] The Cancer Imaging Archive (TCIA), “The Lung Image Database Consortium (LIDC-IDRI),” Public Dataset Repository, 2023. [Online]. Available: https://www.cancerimagingarchive.net
[19] N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” Journal of Machine Learning Research, vol. 15, no. 1, pp. 1929–1958, 2014.
[20] World Health Organization (WHO), “Global Cancer Observatory: Lung cancer fact sheet,” International Agency for Research on Cancer (IARC), 2025. [Online]. Available: https://gco.iarc.fr

BIOGRAPHIES

<table>
<tr><td></td><td>Name: Dr. VIDHYA K
Assoc.Prof , Dept. of ISE & AD
East West Institute of Technology,
Bengaluru-560091
Email id: vidhyak@ewit.edu.in
Contact No.: 8095558259</td></tr>
<tr><td></td><td>Name : MONISHA K
UG Student , Dept. of AD
East West Institute of Technology,
Bengaluru-560091
USN No.: 1EW22AD024
Email id: monishakkumar2004@gmail.com
Mobile No.: 9742728268</td></tr>
</table>

	Name : POOJASHREE A UG Student , Dept. of AD East West Institute of Technology, Bengaluru-560091 USN No. : 1EW22AD029 Email id: ppoojashreeanand@gmail.com Mobile No.: 7676581075
	Name: PRAJWAL K T UG Student , Dept. of AD East West Institute of Technology, Bengaluru-560091 USN No.: 1EW22AD031 Email id: prajwalkt2004@gmail.com Mobile No.: 7975971689
	Name: THANUSHA K M UG Student , Dept. of AD East West Institute of Technology, Bengaluru-560091 USN No.:1EW22AD048 Email id: thanushagowda081@gmail.com Mobile No.: 8971155875



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)